# SuRGE Workshop, Summer 2019
# Rat Genome Database

## Identifying a genomic region using RGD tools

- Literature search (OntoMate)
  - OntoMate is an ontology-driven, concept-based literature search engine developed at RGD as an alternative for the basic PubMed search engine (http://www.ncbi.nlm.nih.gov/pubmed).
  - OntoMate tags abstracts with gene names, gene mutations, organism names and terms from the ontologies/vocabularies used at RGD.
  - For more about ontologies, see RGD's "Introduction to Biomedical Ontologies" video tutorial series at https://rgd.mcw.edu/wg/home/the-introduction-to-biomedical-ontologies-video-series/
  - OntoMate can be accessed from the link in the RGD header (e.g. on the home page https://rgd.mcw.edu), or using this URL: https://rgd.mcw.edu/QueryBuilder/
- Ontology Search
  - The ontology search and browser page is accessible via the "Ontologies" button on the RGD home page or by going to https://rgd.mcw.edu/rgdweb/ontology/search.html
  - Search returns the list of ontologies that match your search criteria with the count of the matching terms in the left margin, and a table of all of the matching terms. Choose an ontology from the list on the left to only display terms from that ontology in the result table.
  - The leaf icon links to the ontology browser. The "A" icon links to the ontology report page listing all genes, etc. annotated to that term or any of its children. The Ontology Report page can also be accessed by clicking the term name in the search result table.
  - Clicking the name of an ontology on the ontology search page (https://rgd.mcw.edu/rgdweb/ontology/search.html) takes you to the ontology browser for that vocabulary. The top level term will be displayed in the center panel of the browser. Browse down the tree by clicking terms in the "Child terms" panel on the right. Browse up the tree by clicking terms in the "Parent Terms" panel on the left.
  - If you sign up for "MyRGD" using the link in the top right corner of the page, you can add a "watcher" to any term in the ontology to receive notification when a new annotation to that term has been assigned. Watchers are also available for gene, QTLs, strains, etc—anywhere a binocular icon is displayed—to receive notification of an update to the record, or a new annotation added for that record.
  - The Ontology Report page shows the data objects annotated to a term and/or its children, with position information for each where that is available. A display of the GViewer tool shows the location of each object relative to a karyotype view of the chromosomes for that species allowing users to see areas where there is a concentration of annotated objects. Click on the chromosome to open the zoom panel. Drag the slider to the desired location. The position of the region covered by the slider is displayed.
- Find a region based on orthology to another species
  - RGD's new Gene and Ortholog Location Finder (GOLF, https://rgd.mcw.edu/rgdweb/ortholog/start.html) gives the orthologs in a second species for a list of genes in an input species.
  - Enter a list of gene symbols or a genomic position of the region of interest in one species, for instance genes or a region identified in a human GWAS study. Select the input species with the information about the assembly for which you would like positions. Select an output species with the information about the assembly for which you would like the list of orthologous genes with their positions.
  - The tool produces a table of the input gene with their positions, and the orthologs in the other species with the corresponding positions. The list can be downloaded into an Excel spreadsheet. Gene symbols in the list link to the corresponding gene report pages at RGD.
  - Output is not currently sorted in any way. Download the list and sort it by position to see blocks where the gene list and the order of those genes correspond between the two species.

## Interrogating a genomic region using RGD tools

- Search for Orthologs (GOLF)
  - The Gene and Ortholog Location Finder (GOLF) is located at https://rgd.mcw.edu/rgdweb/ortholog/start.html
  - For more information on using GOLF to find orthologs for your gene list or region of interest, see above.
- View the Neighborhood (JBrowse)
  - The JBrowse genome browser for rat Rnor 6.0 is located at https://rgd.mcw.edu/jbrowse/?data=data_rgd6&tracks=ARGD_curated_genes&highlight=
  - For a list of all of the genome browsers available at RGD click "Genome" in the header of the rat Rnor 6.0 JBrowse, or go to https://rgd.mcw.edu/jbrowse/.
  - You can select what tracks to view (or turn off) using the list of "Available Tracks" in the left margin of the tool.
  - Click on a category to display the sub-categories and/or specific tracks under it.
    - To view all genes, QTLs or strains in a region, select tracks in the Gene Models, QTLs or Strains categories, respectively.
    - To limit results to only genes, QTLs or strains that have been annotated to terms in specific disease categories, choose "Disease Related Tracks".
    - To limit your results to only genes that interact with chemicals in a specific category, choose "Gene-Chemical Interaction Tracks".
    - To view the reference genomic sequence, select "Reference Data".
    - The "Variants" category includes sub-categories
      - "DbSNPs" for variant records downloaded from dbSNP
      - "Micro Satellite Markers" for RGD's SSLP records
      - "Strain Specific Variants" for all of the variants from WGS of rat strains, and
      - "Damaging Variants" which shows only the variants from WGS of rat strains that have been predicted to be either possibly or probably damaging by Polyphen.
  - For more information on using JBrowse, see our help documentation at https://rgd.mcw.edu/wg/help3/tools/rgd-genome-browsers/the-rat-jbrowse-genome-browser/
- Find Strain-Specific Variants (Variant Visualizer)
  - The Variant Visualizer tool can be accessed from the Analysis & Visualization menu dropdown, or the Analysis & Visualization page at https://rgd.mcw.edu/wg/tool-menu/, or by going directly to https://rgd.mcw.edu/rgdweb/front/config.html.
  - Variation data is available for rat strains on all three major genomic assemblies. ClinVar clinical variants are also available for human build 37 and 38. The default choice is rat Rnor 6.0 assembly.
  - You can start searching for variants by selecting strains, or by specifying a position or a gene list.
  - On the strain selection page, select any of the available strains using the +/- toggles, or select all of the strains by clicking "Select All". The sample name consists of the specific strain/substrain symbol with an abbreviation for the institution at which the secondary analysis of the sequence and the variant calling was done in parentheses.
    - Three options are provided for specifying the genes or region to query: "Limit by Genomic Position" to enter chromosome, start and stop coordinates or the symbols for two genes or markers to bound the region of interest.
    - "Search by Function" to go to the OLGA tool to retrieve a list of genes based on functional (ontology) annotations. See below for more information about OLGA.
    - "Enter a Gene List" to paste in a list of gene symbols, either comma separated or one per line.
  - The "Select Sequence Annotation" page gives options for limiting your results based on variant attributes, sequencing call statistics and/or variant consequences, including limiting results to variants predicted to be possibly or probably damaging by the Polyphen software.
  - If your input is a list of genes, or the query returns a large number of variants for a region, the results will be shown first as a "Variant Distribution", i.e. a count of the number of variants found

for each gene and/or intergenic region in the output.  Click a gene symbol or region to view the variants for that gene/region.

- Clicking a specific variant in the results heatmap displays a popup window with details about the variant, its sequencing call statistics such as read depth, position relative to all overlapping transcripts and the variant consequences for each transcript where that data is available.

## Examining your favorite gene(s) using RGD tools

- Gene Report Page
  - Find your gene of interest by entering the symbol, name or other keywords into the general search box at the top of any RGD page. Or most tools provide links directly to gene report pages.
  - RGD's faceted search results page lets you narrow your results based on species, object type, etc.
  - Click anywhere on the row in the table to access the report page for that gene or other object.
  - The top of the gene report page has general information about the gene as well as information about orthologs, and about genetic models for that gene where applicable.
  - The position section gives position information for multiple genomic assemblies, as well as positions on genetic and RH maps where that data exists. Links are provided to multiple genome browsers.
  - The "Model" pane is a live, interactive snippet of JBrowse embedded in the gene page. You can scroll right or left to view the genes on either side of the one you are looking at, or click "Full-screen view" to access the full JBrowse at that gene location.
  - Ontology annotations are the major source of information about a gene. RGD manually annotates Disease, Gene Ontology, Pathway and Phenotype information from the literature, and imports annotations for Gene-Chemical Interactions (i.e. ChEBI) from the Comparative Toxicogenomics Database.
  - The default view is a summary of the terms and their "evidence codes" associated with the gene. Click "Toggle Annotation Detail/Summary View" at the top of the Annotation section to expand the view to show more details about the annotation, including a link to the reference from which the annotation was made.
  - Two other sections which might be of particular interest to researcher are the "miRNA Target Status" and "Damaging Variants" sections.
    - For genes that are targets of microRNA regulation, miRNA Target Status shows the miRNAs confirmed and/or predicted to regulate expression of that gene.
    - For miRNAs, miRNA Target Status lists the genes which are confirmed and/or predicted to be regulated by that miRNA.
    - Where the data is available, the "Damaging Variants" section lists the locations and the corresponding strains that have variants in that gene that are predicted to be possibly or probably damaging by Polyphen.
- Online List Generator & Analyzer (OLGA)
  - OLGA is RGD's "Advanced Search" and bulk search.
  - OLGA can be accessed by using the link beside the general search box, the links in the Analysis & Visualization dropdown or page, or directly at https://rgd.mcw.edu/rgdweb/generator/list.html.
  - Various criteria can be used to create lists of genes. Search for genes based on annotations to any of the ontologies used at RGD. Enter a genomic region using chromosome, start and stop coordinates. Enter a QTL symbol to retrieve the list of genes that overlap that QTL. Or enter a list of gene symbols into the input box ("Symbol List").
  - Multiple lists can be combined in various ways, including adding the lists together to get a complete non-redundant list of the genes in all of the input lists, subtracting one list from another to retrieve the genes in the first list that do not appear in the second, or intersecting the lists to obtain the genes that overlap in the two lists.
  - Once the desired final result is obtained, use "Analyze Result Set" to access the toolbox. Selecting a tool in the toolbox automatically sends the list of genes from the final Result Set to that tool for analysis. In addition to the Genome Viewer, Variant Visualizer, and GOLF mentioned previously, the following tools are available in the toolbox. An option is also provided to download the list of genes as an Excel file to keep for your records.
- Gene Annotator Tool
  - The Gene Annotator (GA) Tool is available from the RGD Toolbox, by using the links in the Analysis & Visualization dropdown or page, or at https://rgd.mcw.edu/rgdweb/ga/start.jsp.

- Select a species, then enter a list of gene symbols or other identifiers, or specify a genomic region.
- Select your desired output, including ontologies, external database (xdb) links and/or species to obtain orthologs. The default is for all of the information to be included.
- The GA Tool has three major sections:
  - Annotations: shows the orthologs and xdb links for each gene and the ontology annotations for the gene and all of its orthologs.
  - Annotation Distribution: Lists the terms from each ontology in descending order of the proportion of genes in the list with annotations to that term. For each term, the genes from the list that are annotated to the term or any of its children are listed and those subsets of the original list can be resubmitted to the GA Tool for analysis. Select multiple terms using the check boxes to the right of the term to obtain and analyze the list of genes annotated to all of the selected terms ("Cross Analysis"). Note that this tool does not do statistical enrichment analysis. For enrichment, see MOET below.
  - Comparison Heat Map: Use the comparison heat map to view the number and list of genes annotated to terms in any two ontologies or any two branches of the same ontology. Each column and row title is a link which opens the list of direct children of that term, making it easy to navigate down the tree to see the overlap of more specific annotations.

- Protein-Protein Interactions (InterViewer)
  - The InterViewer Tool is available from the RGD Toolbox, by using the links in the Analysis & Visualization dropdown or page, or at https://rgd.mcw.edu/rgdweb/cytoscape/query.html.
  - InterViewer takes lists of UniProtKB protein IDs, gene symbols or gene RGD IDs.
  - InterViewer provides an interactive view of the proteins ("nodes") and their interactions ("edges") in a Cytoscape-based viewer. Select a node or an edge to see more information about that protein or interaction. Multiple selections are provided for the layout of the interactions. Nodes can be dragged to make the interactions easier to see.
  - Selecting a node highlights that protein and all of its interactions and interacting partners.
  - A downloadable table of all of the interactions for the input set of genes/proteins is provided.
  - If the query returns too many interactions to be usefully visualized, the tool provides the option to download the full set of results.

- Multi Ontology Enrichment Tool (MOET)
  - MOET, one of RGD's newest tools, is available from the RGD Toolbox, by using the links in the Analysis & Visualization dropdown or page, or at https://rgd.mcw.edu/rgdweb/enrichment/start.html
  - Select a species and an ontology to start, and enter a list of gene symbols or other identifiers. For identifiers, you will need to specify the type of ID you are inputting. The default is rat as the species, disease as the output ontology and gene symbols as the input type. Alternatively, enter the chromosome, start and stop for a genomic region to perform enrichment for all of the genes in the region.
  - The result page shows the list of enriched terms, i.e. those which are statistically more highly represented in annotations for your gene list than for the list of all gene for your selected species. A graph is also provided that shows the number of genes and the respective p-values for each of the enriched terms.
  - Use the options at the top of the page to select a different species and/or different ontology. There is also an option to view the enrichment for a given ontology across all species. The gene lists for other species are created using the ortholog assignments for the input gene list.

For more information about RGD tools, please see the help pages at https://rgd.mcw.edu/wg/help3/ and/or the video tutorials at https://rgd.mcw.edu/wg/home/rgd_rat_community_videos/rgd-tool-and-website-videos/. Help documentation and video tutorials are coming soon for the newer tools such as MOET and GOLF. If you have any questions, or would like to request that a video be made on a specific topic, please contact us at https://rgd.mcw.edu/wg/contact/.